

Fig. S6. Number of genes with motifs identified by Xie et al., 2005 in the different core promoter categories (1).

Txn Factor	Motif Index	Position Bias	Motif	TATA + INR				Flexible				TATA only		INR only		None		TATA-plus	TATA-less	Ratio %TATA-less / %TATA-plus
				Module		Flexible		TATA only		INR only		None		None		(2434 genes) %	(7837 genes) %			
				No. genes (211 total)	%	No. genes (1358 total)	%	No. genes (865 total)	%	No. genes (3107 total)	%	No. genes (4730 total)	%	No. genes (4730 total)	%	(2434 genes) %	(7837 genes) %			
NRF-1	M1	-62	RCGCANGCGY	0	0	3	0.22	7	0.81	33	1.06	52	1.1	0.4	1.1	2.8*				
MYC	M2	-62	CACGTG	7	3.32	36	2.65	34	3.93	159	5.12	296	6.26	3.2	5.8	1.8				
ELK-1	M3	-24	SCGGAAGY	3	1.42	46	3.39	35	4.05	253	8.14	517	10.9	3.5	9.8	2.8				
-	M4	-89	ACTAYRNNNNCCR	1	0.47	16	1.18	3	0.35	50	1.61	47	0.99	0.8	1.2	1.5				
NF-Y	M5	-63	GATTGGY	8	3.79	36	2.65	45	5.2	119	3.83	196	4.14	3.7	4	1.1				
Sp1	M6	-63	GGCGGGR	13	6.16	89	6.55	160	18.5	634	20.4	1514	32	10.8	27.4	2.5				
AP-1	M7	-	TGANTCA	9	4.27	65	4.79	29	3.35	93	2.99	88	1.86	4.2	2.3	0.5				
-	M8	-62	TMTGCCANR	0	0	8	0.59	2	0.23	34	1.09	38	0.8	0.4	0.9	2.3				
ATF3	M9	-66	TGAYRTCA	5	2.37	35	2.58	23	2.66	101	3.25	134	2.83	2.6	3	1.2				
YY1	M10	-	GCCATNTTG(INR)	2	0.95	19	1.4	2	0.23	86	2.77	38	0.8	0.9	1.6	1.8				
GABP	M11	-23	MGGAAGTG	3	1.42	44	3.24	25	2.89	139	4.47	280	5.92	3	5.3	1.8				
ATF3	M14	-22	TGACGTCA	2	0.95	13	0.96	16	1.85	60	1.93	107	2.26	1.3	2.1	1.6				
C-ETS-2	M16	-24	RYTTCTG	12	5.69	46	3.39	22	2.54	151	4.86	135	2.85	3.3	3.6	1.1				
-	M21	-33	GGAANCAGAANY	0	0	2	0.15	8	0.92	15	0.48	35	0.74	0.4	0.6	1.5				
-	M22	-17	TGCGCANK	2	0.95	33	2.43	34	3.93	138	4.44	505	10.7	2.8	8.2	2.9				
-	M30	-31	YGCGYRCGC	3	1.42	20	1.47	24	2.77	91	2.93	270	5.71	1.9	4.6	2.4*				
C-REL	M48	-12	GCGNNANTTC	0	0	7	0.52	7	0.81	55	1.77	40	0.85	0.6	1.2	2				

Method:

The -250 to +150 region of the indicated promoter categories were searched for the 11 most conserved motifs (M1-M11) and 6 other motifs (M14, M16, M21, M22, M30 and M48) with a position bias between -40 and +1 from the 50 most highly conserved motifs across several mammalian species identified in Table 1 by Xie et al., 2005.

Legend:

Transcription (Txn) Factor, Motif Index, Postion Bias and Motif are from Xie et al., 2005.

Module: Genes with a TATA box and an INR in which the TATA box is 15 to 30 nt upstream of the INR, within the -80 to +80 region.

Flexible: Genes with a TATA box and an INR in any configuration in the -80 to +80 region.

"TATA only," "INR only" and "None" are as described in Fig. 5 (main text).

TATA-plus: all genes with a TATA-532 sequence ("TATA+INR" plus "TATA only").

TATA-less: all genes lacking a TATA-532 sequence ("INR only" plus "None").

The total number of genes in each promoter category is given in parentheses.

All TATA's are TATA-532 as described in Fig. 1 (main text).

Highlight, motifs chosen for further analysis in Fig. 6 (main text) due to an enrichment in TATA-less promoters (last column).

* The M1 motif was not pursued due to a low number of genes. M30 was not pursued due to substantial overlap in the consensus sequence with M22.

Note: percentages given here refer to the percent of genes within a given promoter category (e.g., TATA only) that contain a given motif (e.g., M6).

In contrast, the percentages in Fig. 6A (main text) refer to the percent of genes with a given motif that fall within a given promoter category.

Reference:

- (1) Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., Kellis, M. (2005)
Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals.Nature 434: 338-345.